

Overview

Motivation

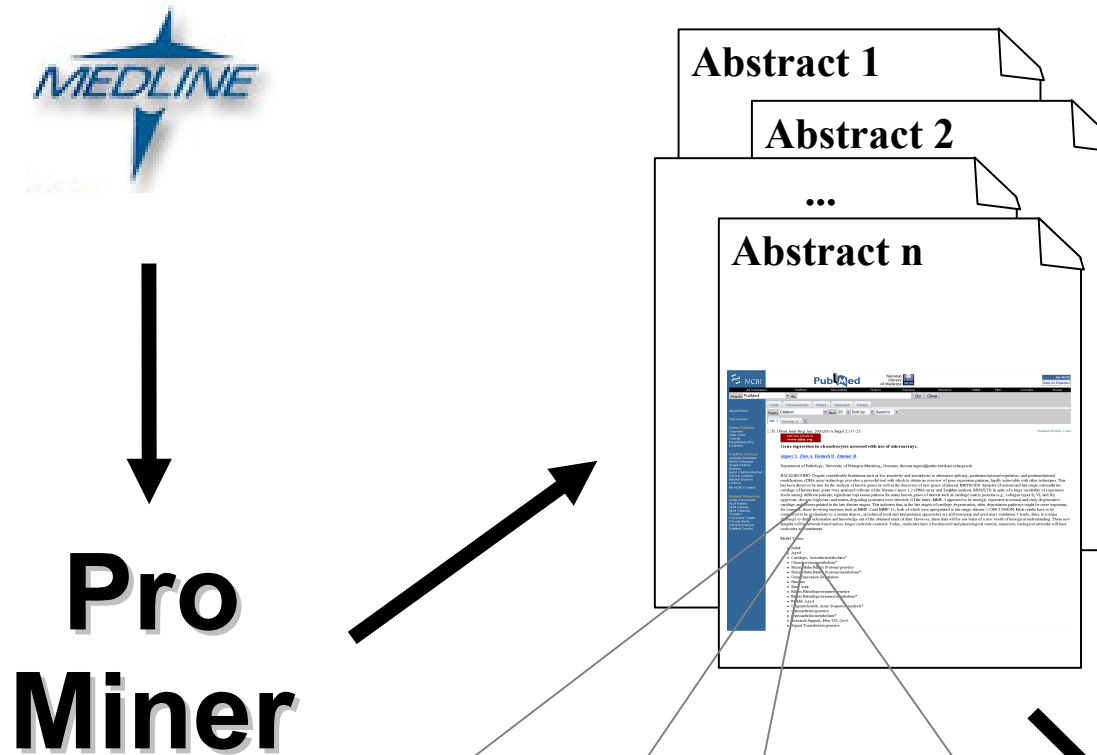
- Standard gene expression analysis
- Based on statistical expression analysis alone
- Results difficult to interpret without using biological background knowledge
- Our goals
 - Exploit biological literature
 - Find disease specific gene groups that are differentially expressed and share a coherent literature background
- Application
 - Evaluation of gene expression data from osteoarthritis (OA) patients
 - 83 cDNA arrays comprising 7467 spots, classified into normal, early degenerative cartilage, peripheral and central OA samples

Osteoarthritis related terms

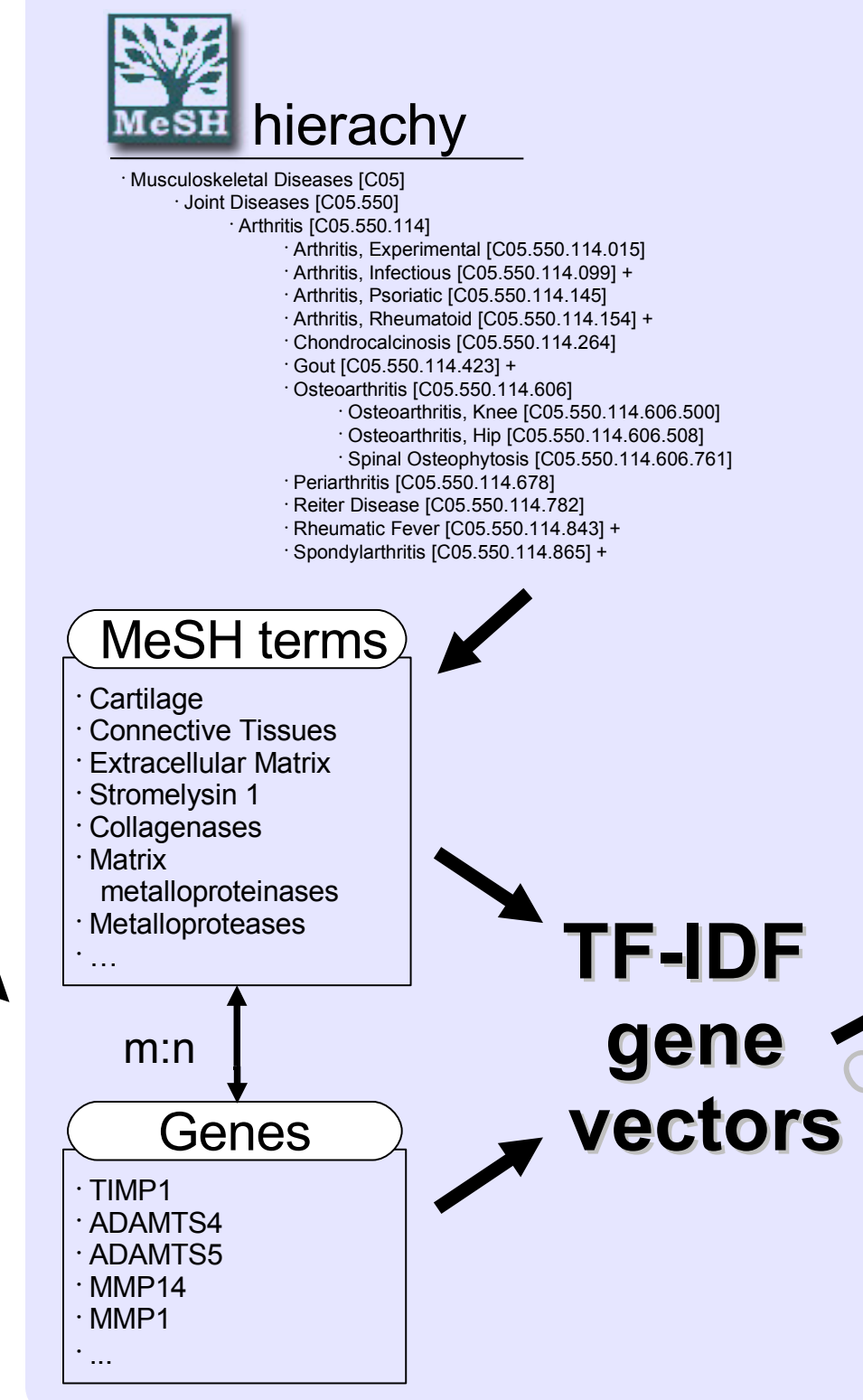
Hypertrophic pulmonary osteoarthropathy
Marie Bamberg Disease
Bamberg-Marie Disease
HYPEROSTROTIC ARTHRITIS
Osteoarthritis
ARTHRITIS, SENESCENT
Cervical or osteoporosis spondylarthritidis
Cervical spondylarthritidis
Cervical Spondylitis
Cervical Spondylosis
Osteoarthritis Deformans
Kocher-Back Diseases
Chondrocytes
Synovial cartilage
Synovial fluid

Human gene synonym list

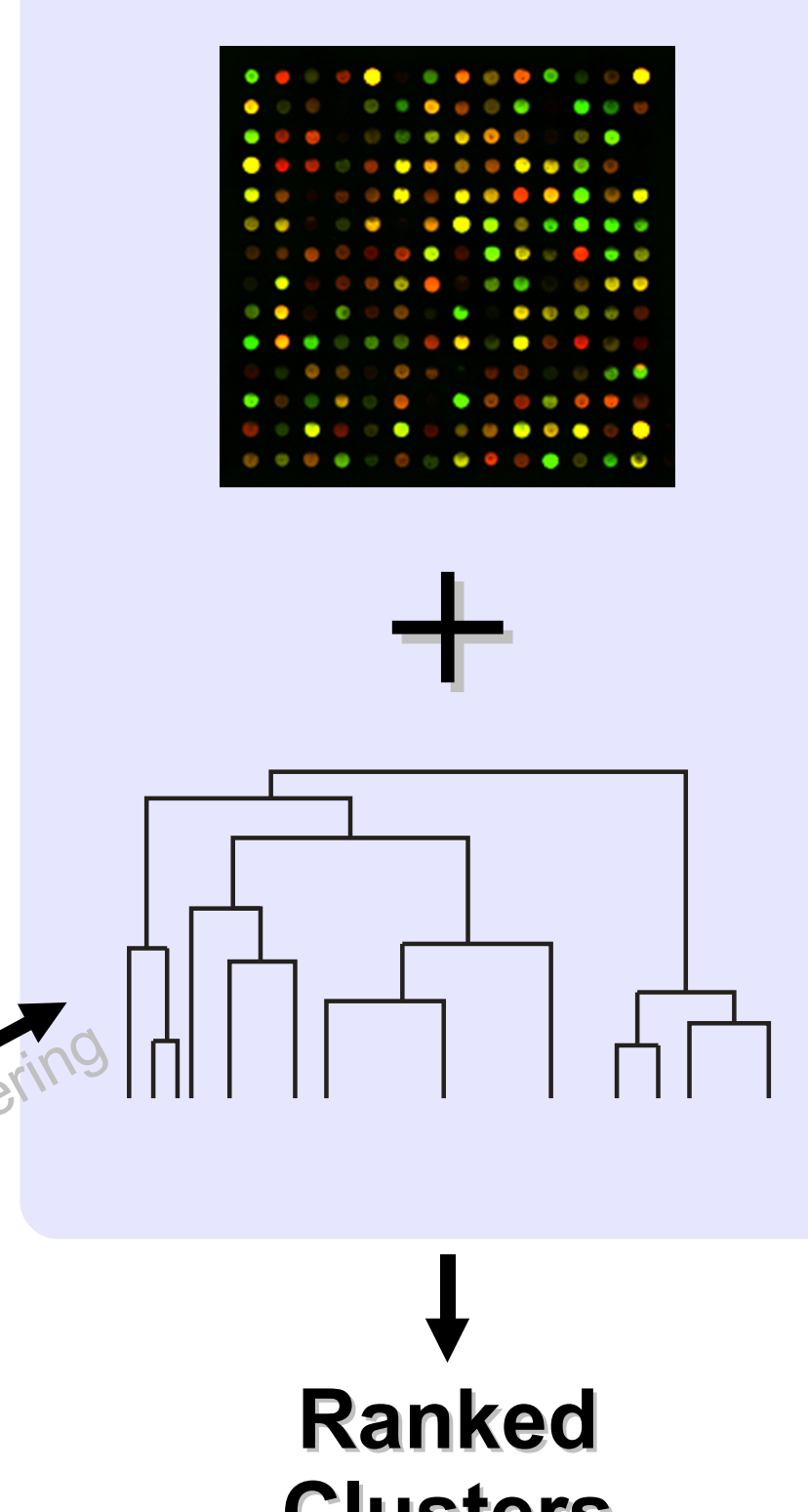
MAPK8IP2
JNK-interacting protein 2
Mitogen-activated protein kinase 8-interacting protein 2
Islet brain-2
MAP kinase 8 IP 2
COL17A1
150 kDa bullous pemphigoid antigen 2
BP 150
BS-150/23.2
collagen, type XVII, alpha-1
MMP15
MT 2 matrix metalloproteinase
matrix metalloproteinase 15
Membrane-type-2 MMP



Feature Construction



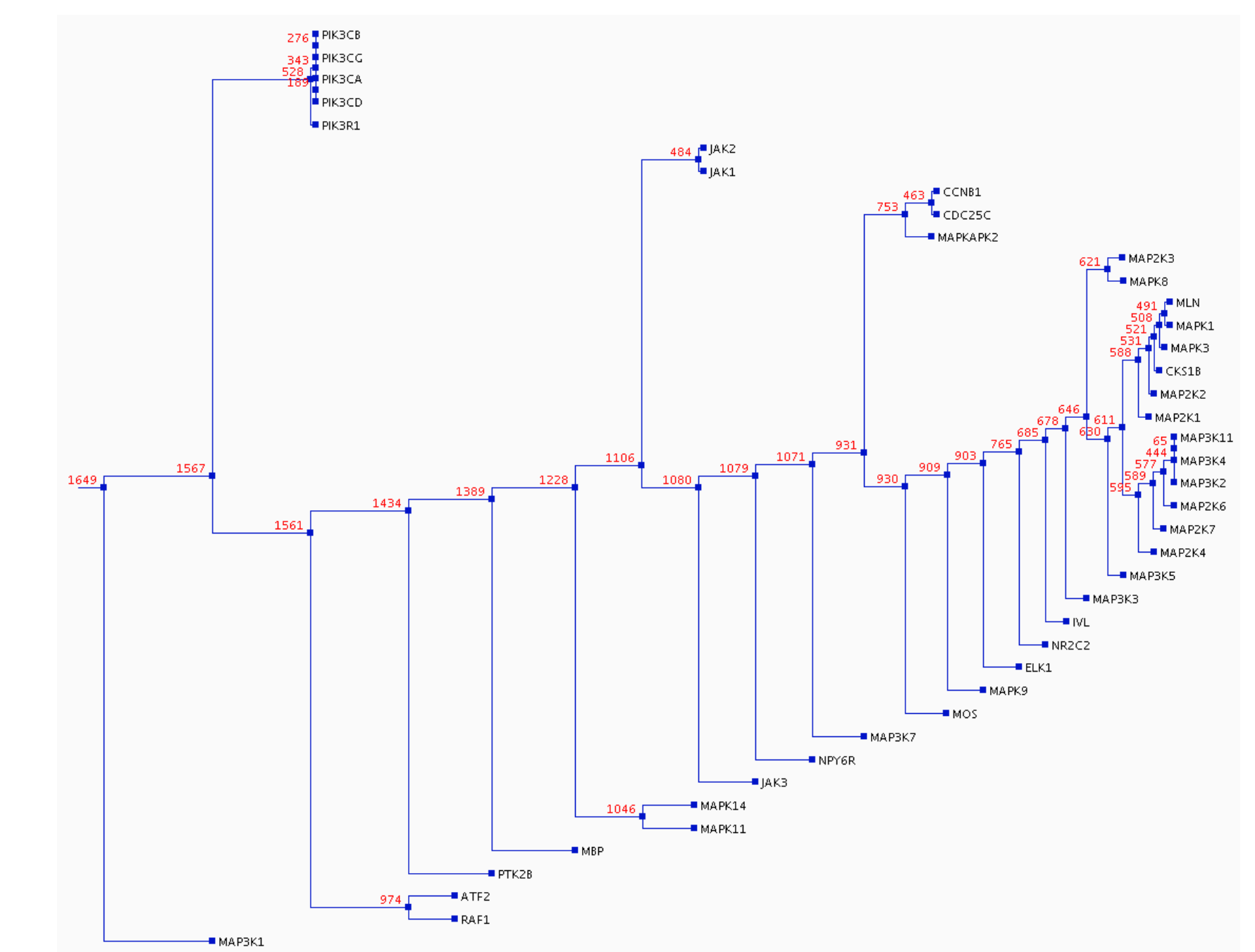
Cluster Ranking



Feature Construction

Identification of disease-related gene clusters

- Initial gene set
 - We mine MEDLINE for abstracts with ProMiner (Hanisch et al. 2003) to find co-occurrences of human genes and OA related terms
- Feature Vector Construction
 - Identification of features by exploiting the MeSH keyword hierarchy
 - Each gene feature vector consists of MeSH terms
 - Feature values are quantified using a modified version of TF-IDF ($TF \times DF^{-1}$)
 - TF: How often MeSH term i is associated with gene j
 - DF: How many genes are associated with MeSH term i
- Hierarchical clustering approach
 - Genes belonging to one cluster share a coherent pattern of MeSH term combinations
 - Normalized scalar product is used as distance measure



Subtree (i.e. possible interesting gene group)

PMID- 15880812
MH - Chondrocytes
MH - Chondrosarcoma/metabolism/pathology
MH - Synovial fluid
MH - Endopeptidases/biosynthesis
MH - Extracellular Signal-Regulated MAP Kinases
MH - Cartilage
MH - Interleukin-1/pharmacology/physiology
MH - Metalloendopeptidases/biosynthesis
MH - Osteoarthritis knee
SO - Arthritis Rheum 2005 May;52(5):1451-60.

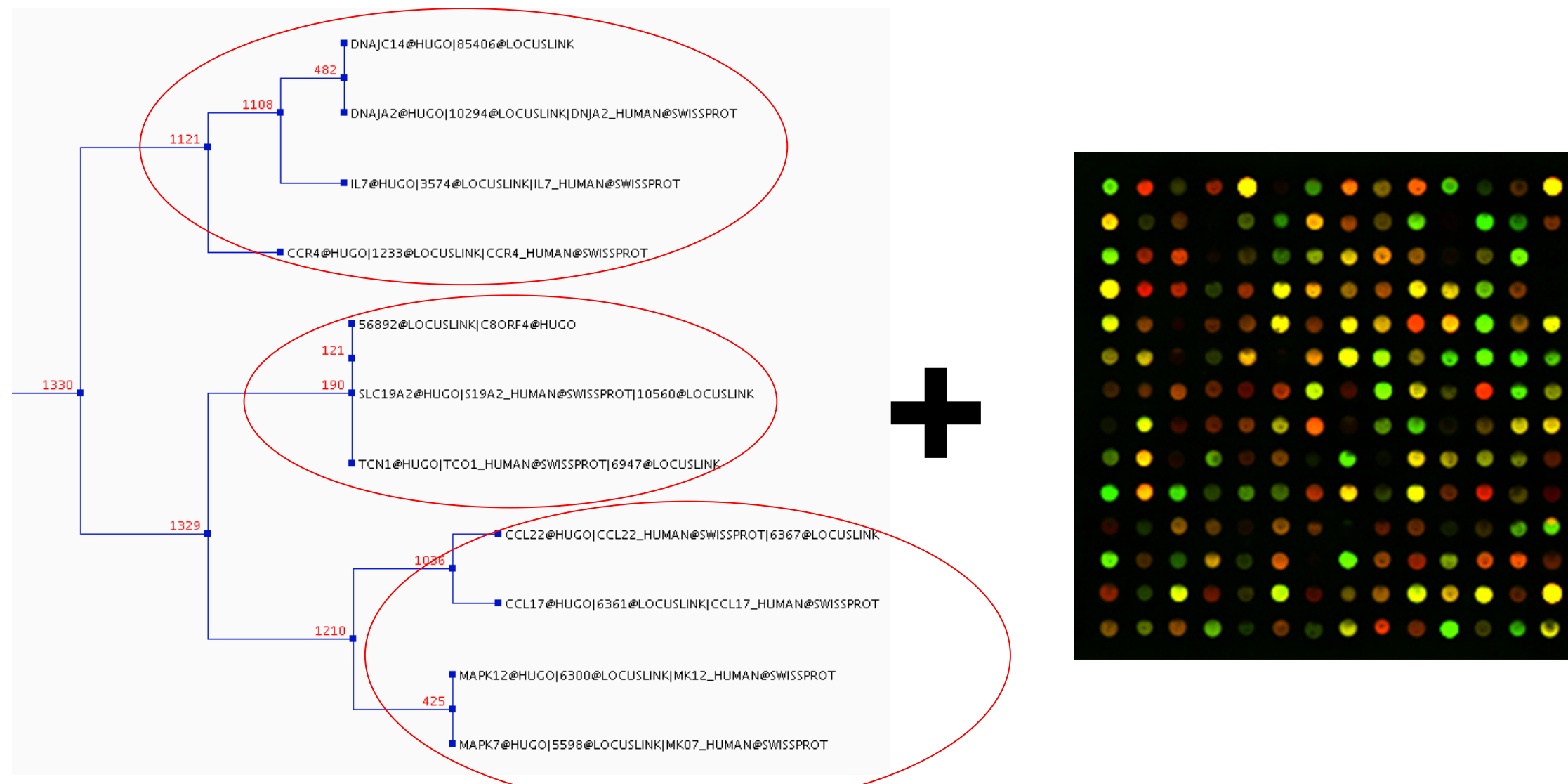
	Osteo arthritis knee	Synovial fluid	Chondro sarcoma	Cartilage	Chondro cytes
COL5A1	0.7	0	0.23	0.65	0
COL5A2	0.8	0	0.37	0.25	0
THSD3	0.03	0	0.09	0.01	0

Construction of gene specific feature vectors using abstract associated MeSH terms

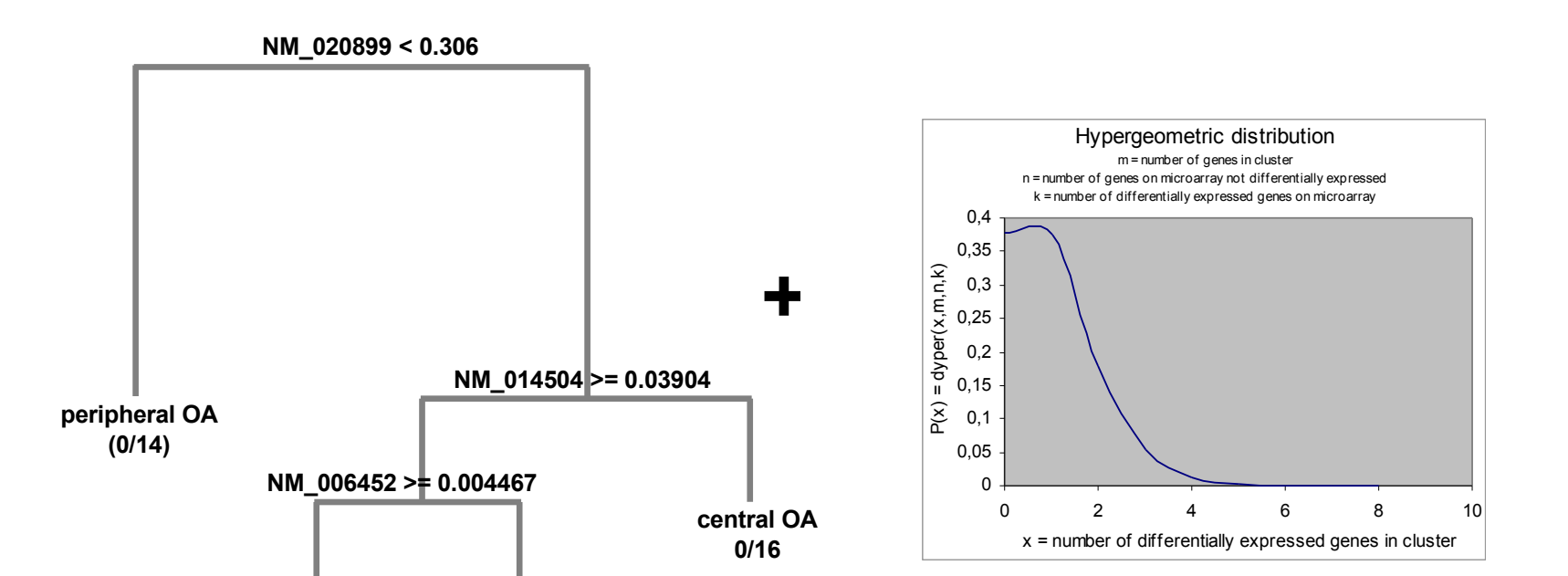
Cluster Ranking

Cluster Ranking

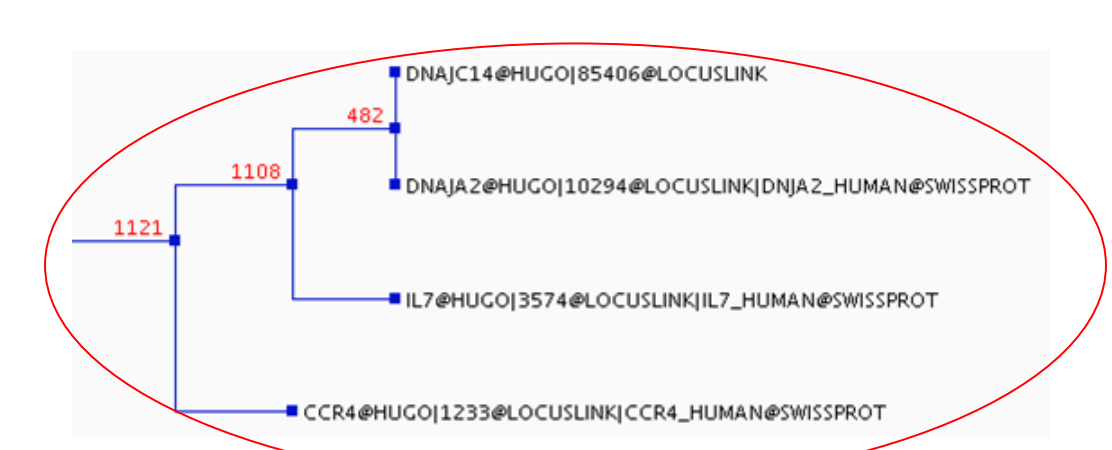
- Goal
 - Identify literature clusters relevant to the given disease context with the help of gene expression data
- Cluster ranking to assess the importance of clusters
 - Decision tree analysis (DT)
 - Verified using ten fold cross validation
 - Accuracy is used for ranking
 - Provides a measure of how well genes of a cluster are suited to distinguish between different OA stages
 - Overrepresentation analysis (ORA)
 - Clusters are ranked corresponding to enrichment of differentially expressed genes
- Overall cluster ranking using weighted rank sum from DT + ORA



Starting from our gene clusters obtained by literature analysis we try to identify clusters relevant to the OA context with the help of gene expression data



A weighted combination of decision tree classifier and overrepresentation analysis is used for cluster selection



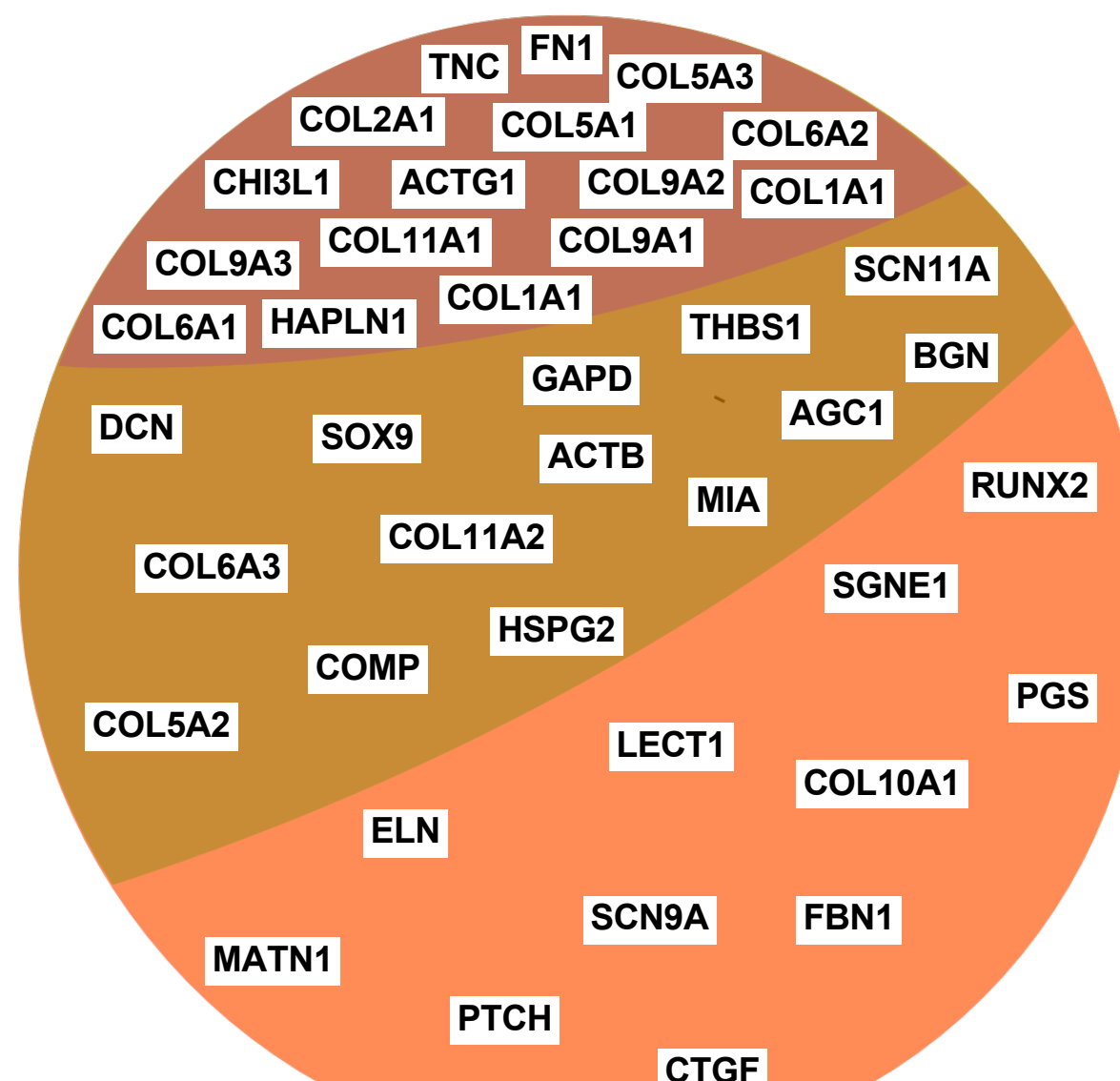
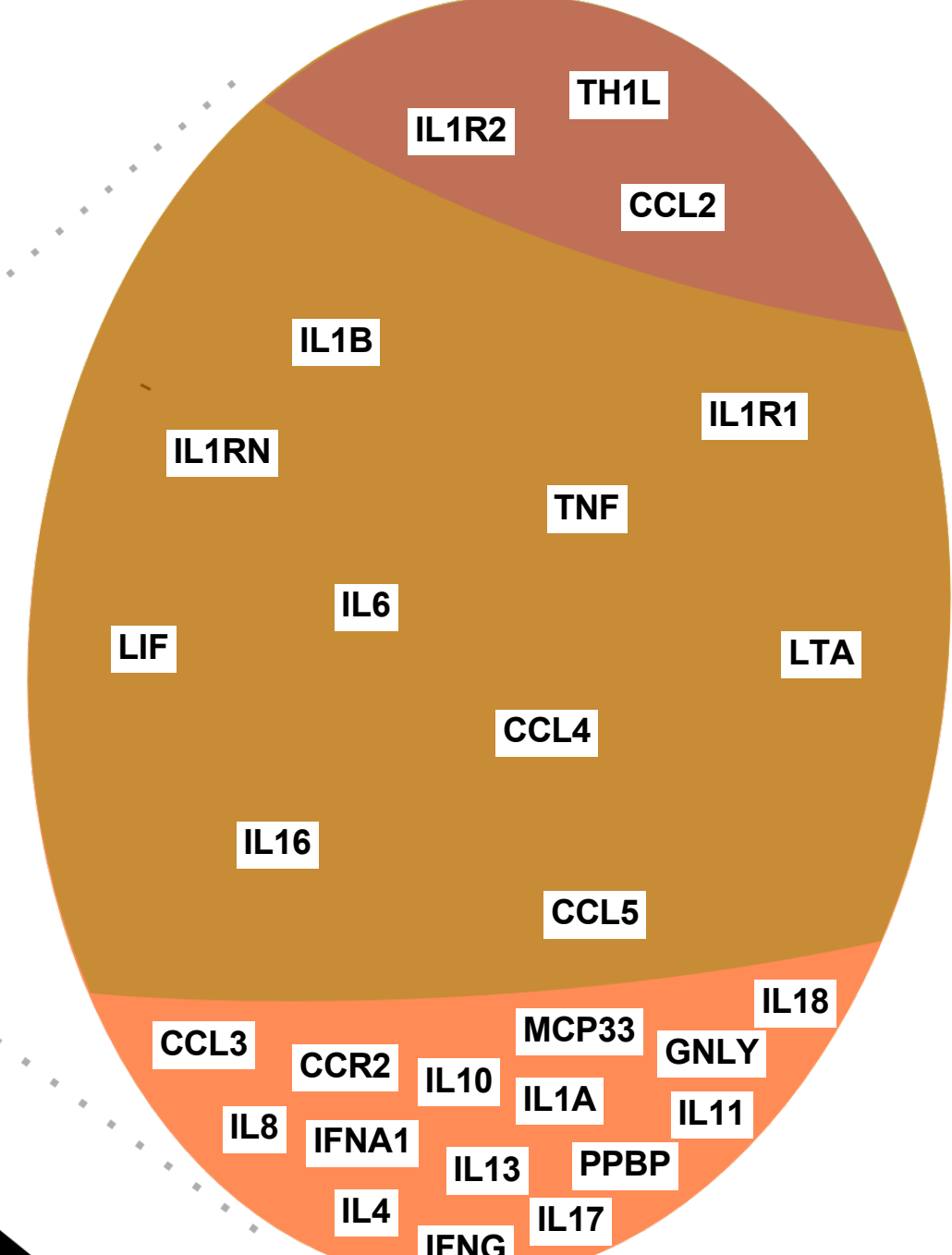
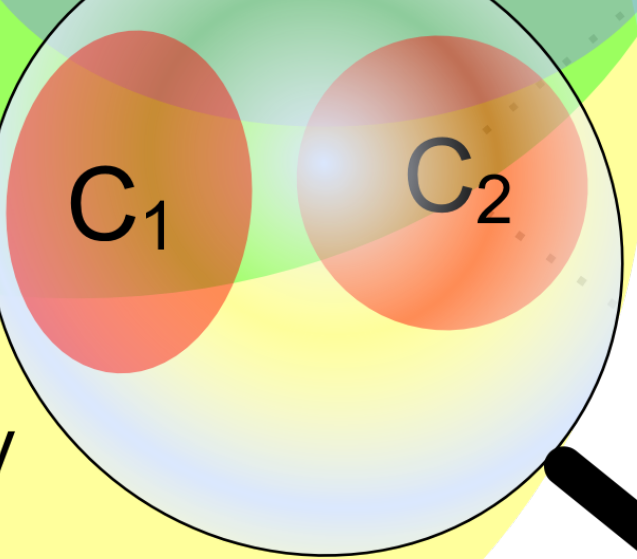
Resulting interesting gene cluster

Results

genes on microarray

differentially expressed genes

genes found by text mining



Conclusion

- Results
 - Top ranking clusters contain OA relevant genes (collagens, MMPs, ADAMs and interleukins)
- Features
 - Clusters might predict additional genes not measured on the chip
 - Starting point for improved chip design
 - Insights into underlying biological mechanisms
- Outlook
 - Optimize clusters by iterative procedure
 - Report publications relevant to clusters
 - More in-depth evaluation of clusters
 - Incorporate additional features in feature vectors